Accepted manuscript

Validation of a life-logging wearable camera method and the 24-hour diet recall method for assessing maternal and child dietary diversity

Andrea L. S. Bulungu[1], Luigi Palla[1], Jan Priebe[2], Lora Forsythe[2], Pamela Katic[2], Gwen Varley[2], Bernice D. Galinda[1], Nakimuli Sarah[1], Joweria Nambooze[3], Kate Wellard[2] and Elaine. L. Ferguson[1]

[1]*London School of Hygiene and Tropical Medicine, Keppel St, London WC1E 7HT, United Kingdom*

[2]*Natural Resources Institute (NRI), University of Greenwich, Medway Campus, Central Avenue, Chatham Maritime ME4 4TB, United Kingdom*

[3]*Africa Innovations Institute (AfrII), Plot 1544, Koire Close, Off Old Kiira Road, Bukoto, Uganda*

Author responsible for correspondence: Andrea L. S. Bulungu; London School of Hygiene and Tropical Medicine, Keppel St, London WC1E 7HT, UK; andrealspray@gmail.com; +1 (202) 250-4399

Short title: Validation of a camera method for assessing diet

Keywords: validation studies; diet assessment; wearable camera; Bland Altman;

This peer-reviewed article has been accepted for publication but not yet copyedited or typeset, and so may be subject to change during the production process. The article is considered published and may be cited using its DOI
10.1017/S0007114520003530
The British Journal of Nutrition is published by Cambridge University Press on behalf of The Nutrition Society

Abstract

Accurate and timely data are essential for identifying populations at risk for undernutrition due to poor quality diets, for implementing appropriate interventions, and evaluating change. Life-logging wearable cameras (LLWC) have been used to prospectively capture food/beverage consumed by adults in high income countries. This study aimed to evaluate the concurrent criterion validity, for assessing maternal and child dietary diversity scores (DDS), of a LLWC-based image-assisted recall (IAR) and 24-hour recall (24HR). Direct observation was the criterion method. Food/beverage consumption of rural Eastern Ugandan mothers and their 12-23-month-old child (n=211) were assessed, for the same day for each method, and the IAR and 24HR DDS compared with the weighed food record (WFR) DDS using the Bland-Altman limits of agreement (LOA) method of analysis and Cohen's kappa. The relative bias was low for the 24HR (-0.1801 for mothers; -0.1358 for children) and the IAR (-0.1227 for mothers; -0.1104 for children), but the LOA were wide (-1.6615 to 1.3012 and -1.6883 to 1.4167 for mothers and children via 24HR, respectively; -2.1322 to 1.8868 and -1.7130 to 1.4921 for mothers and children via IAR, respectively). Cohen's kappa, for DDS via 24HR and IAR, was 0.68 and 0.59, respectively, for mothers, and 0.60 and 0.59, respectively, for children. Both the 24HR and IAR provide an accurate estimate of median dietary diversity, for mothers and their young child, but non-differential measurement error would attenuate associations between DDS and outcomes, thereby under-estimating the true associations between DDS–where estimated via 24HR or IAR–and outcomes measured.

Introduction

Globally, undernutrition is the single biggest contributor to child mortality[1]. Although the underlying determinants of undernutrition are complex and interacting, inadequate nutrient intake is an immediate cause[2,3]. The prevalence of undernutrition is decreasing, however there are still 144 million children under-5 years who suffer from chronic malnutrition, more than a third of whom live in Africa[4]. Accurate and timely data are essential for identifying populations at risk for undernutrition due to poor quality diets, for implementing appropriate interventions, and evaluating change[5–8].

Current methods of quantitative dietary assessment are reliable but resource-intensive. To address the need for a rapid, inexpensive, and simple-to-administer method with a low participant burden, reliable population-level food group indicators for measuring diet quality requiring only semi-quantitative dietary data were developed. Cross-country analyses assessing their performance for predicting nutrient adequacy have shown moderate, but variable, associations[9–11]. Two indicators in particular–a dietary diversity score for children under-5 years and a dietary diversity score for women of reproductive age living in low-and middle-income countries–have been validated and are now in widespread use globally[12–14]. Corresponding global minimum dietary diversity standard thresholds have also been validated and are in widespread use[12,13,15]. Assessing dietary diversity using the dietary diversity score or other food group-based indicators is, in comparison, much simpler, and requires only assessing whether any representative foods of each food group in the index was consumed.

"Gold standard" quantitative dietary assessment methods, such as direct observation and repeated weighed food records, are used to accurately and reliably assess the foods consumed as well as the amount of eatable portion of each food consumed by individual consumers. They are seldom used to routinely collect dietary data due to high financial and time costs, participant burden, and requisite expertise[16–18]. Instead, retrospective diet assessment tools using a multiple pass 24-hour "free" recall technique, which entails mothers recalling all foods consumed by their child or themselves on the previous day, are commonly used for estimating food consumption. The accuracy of all food recall methods relies upon the respondent's memory and motivation, as well as the skill and persistence of the interviewer. Misreporting of foods consumed may occur, either unintentionally–for example, for foods that are infrequently consumed–or intentionally, due to interviewer, social desirability, or approval biases[14]. Such errors can result in either underreporting or over-reporting (or both) of food groups defining the dietary diversity score. The 24-hour recall method is also susceptible to other measurement errors[19,20].

There is a long history of using photos to overcome the limitations of traditional food recall methods. "Portion-size estimation aids" involve the use of gradated food photos (representing the range of portion sizes commonly consumed) when collecting semi-quantitative data[21–31]. They are among the earliest approaches to incorporating food photography into diet assessment methods to reduce recall bias and have now been validated in both high and low income countries. Another approach to incorporating food photography (or illustrations) into traditional diet assessment methods is the use of a pictorial food chart to help low-literacy subjects in Sub-Saharan Africa prospectively self-record foods consumed[17,32,33].

In high income country contexts, validation research has also been conducted, in controlled settings, into the use of prospective photos taken by the study subject (often using a mobile phone) during food preparation and/or eating episodes[34–37]. In these studies, information documented in the photos regarding foods consumed, portion sizes, and wastage is later used by researchers to compute nutrient intake. Such "active" photography methods may reduce unintentional errors but do not eliminate the possibility of intentional over- or under-estimation due to social desirability bias. And, although often preferred to traditional methods, active prospective food photo methods require a high level of technical competence among study participants, which further limits its use in large-scale surveys.

New digital media technologies offer opportunities for improving upon traditional 24-hour recall methods of assessing dietary diversity[38–41]. A life-logging wearable camera worn by a study participant can be used to prospectively and passively capture food/beverage consumed, which may provide a more objective method of data collection for assessing women’s and children’s dietary diversity than the 24-hour recall method, with low respondent and interviewer burden. In high income countries, limited validation research on image-assisted recall methods for assessing quantitative dietary data suggest that a life-logging wearable camera can reduce underreporting of energy intake[42,43]. Yet questions remain about their validity, acceptability and feasibility[44,45].

In addition to a range of socio-cultural factors that may affect the acceptability of the life-logging wearable camera in rural low-income country settings, the contexts pose unique technical challenges for wearable cameras, such as lack of electricity with which to illuminate food preparation and consumption at night or indoors and to charge the wearable devices; rugged conditions that expose the camera to water and dirt; and lack of familiarity among participants to digital technology and social media–in particular to first-person photos–which may hamper their interpretation. Furthermore, logistical challenges can be anticipated in assessing dietary diversity for young children. For example, using a life-logging wearable camera attached to the caregiver may not fully capture a child's food intake if the child spends substantial time out of the direct supervision of the primary caregiver (and therefore not in sight of the camera).

This study was undertaken in rural Eastern Uganda to evaluate the concurrent criterion validity, for assessing maternal and child dietary diversity score and minimum dietary diversity, of a life-logging wearable camera-based image-assisted recall method and the 24-hour recall method. Direct observation with weighed food record was the criterion method. Previous studies have evaluated the validity of photo-assisted methods to assess nutrient intake in high income country contexts. No study, to our knowledge, has examined the validity, for estimating dietary diversity score or minimum dietary diversity, of the 24-hour recall method or an image-assisted recall method using a life-logging wearable camera in either free-living or controlled settings.

Methods

Study design

A cross-sectional study of mothers and their child aged 12 to 23 months (n=211) was conducted between January and February (dry season) 2018 in Bugiri and Kamuli Districts, Eastern Region, Uganda. This study was nested within another study designed to examine the impact of a labour-saving technology (a mechanised maize sheller) on women's time for childcare, food preparation and dietary practices.

In our study, food/beverage consumption of mothers and their child were assessed, for the same day, using three concurrent methods: (1) direct observation (15 hours) via weighed food record (WFR), (2) 24-hour recall (24HR), and (3) image-assisted recall (IAR) using a life-logging wearable camera (wearable camera). Data were collected over five consecutive days, following one of two possible patterns (Figure 1). Specifically, for both patterns, on day 1, eligibility was confirmed, a structured questionnaire was administered, and anthropometric data were collected for all participants. Day 1 data were collected at a predefined meeting place in the village. All other data were collected at the participants' home. For half of the study participants, on day 2, food/beverage consumption data were collected using direct observation via WFR and recorded on the wearable camera attached to the mother. On day 3, a 24HR was administered, followed by an IAR using photos captured on day 2 by the wearable camera. On day 4, food/beverage consumption data were again recorded via the wearable camera only (i.e. no observation). On day 5, an IAR was administered using photos captured on day 4 by the wearable camera. The other half of the study participants began with the wearable camera only (i.e., days 2 and 4 were switched) and ended with all three methods (i.e., days 3 and 5 were switched). For all participants, on the 5th day, a final structured questionnaire was also administered. All data collection was performed by trained enumerators. Dietary data collection was distributed across all days of the week to minimise any day-of-the-week effect, and for each mother-child dyad, the enumerator assigned to conduct the direct observation was different from the enumerator assigned to administer the 24HR and IAR.

Ethical approval was obtained from the Uganda National Council for Science and Technology (UNCST) (A24ES), the London School of Hygiene & Tropical Medicine Observational / Interventions Research Ethics Committee (Project ID: 1420), and the University of Greenwich Faculty of Engineering and Science Ethics Committee (Project ID: B0501). The data collection protocols followed the ethical guidelines for life-logging wearable camera research to ensure that privacy of the participants was maintained[38]. Following community sensitization, verbal explanation of the study, and demonstration of the wearable camera, written consent or thumb print was obtained from all mothers who participated in our study.

Participants and sampling

Twenty-two villages were purposefully selected, for this study, of which eleven had access to labour-saving technology and eleven did not. These villages participated in the Sasakawa Global 2000 Uganda (SG2000 Uganda) country program (the local implementing partner for the parent study). The sample size calculation (n= 264; 22 communities, 12 HH per community) was based on requirements of the time allocation study within which this current study was nested. A sample size of 132 per group enabled detection of a 30-minute inter-group difference between women with access to a labour-saving device and households without access to a labour-saving device, assuming a SD of 49 minutes, a design effect of 1.47, 80% power, and p-value of 0.05 and allowing for 10% attrition. This sample size was deemed sufficient for the current validation study, using the Bland-Altman method of analysis[46–49].

The sampling frame, for each village, was a household listing of all mothers with children born between 1/1/2016 and 1/5/2017 inclusive (to recruit children aged 12 to 24 months at the time of data collection). These lists were generated by the SG2000 community-based facilitators. Twelve mother-child dyads in each village were randomly selected to participate in the study; Substitutions were made, as needed, until 12 mother-child dyads who met the inclusion /exclusion criteria were recruited. Mother-child dyads were excluded if the child was less than 12 months or greater than 23 months of age, was not yet eating solid foods on a regular basis, was a multiple-birth child, the mother was unable to communicate in Lusoga, Luganda or English; either the mother or child had a severe disability; the mother was not the biological mother of the child; the mother was a co-wife with another mother selected to participate in the study; or either the mother or child was not available for the duration of the study.

Instruments and protocol

The enumerators administered two structured questionnaires to the mother. The first questionnaire collected information on: household socio-demographics, wealth (adapted from the Uganda 2012 Poverty Probability Index (PPI)), expenditure, and production (adapted from the Abbreviated Women's Empowerment in Agriculture Index (A-WEAI)); knowledge, attitudes, and practices regarding infant and young child feeding and care; factors related to women's empowerment (adapted from the A-WEAI); and the child's health (adapted from Demographic and Health Surveys (DHS)). The second questionnaire was administered at the end of the period of data collection. Response options in this questionnaire included a 4-point Likert for questions related to participants' perceptions of their experiences with each of the three food/beverage data collection methods; and categorical scales for questions related to mobile phone access, ownership and use, and willingness to participate in a future study using 24HR or wearable cameras; in addition to open-ended comments. Duplicate, serial anthropometric measurements of weight and height/length were taken of participating mothers and children.

For the criterion diet assessment method (i.e., WFR), enumerators weighed and recorded all food/beverage consumed by mothers and children, from approximately 06:00 to 21:00 using dietary scales (±1g, Salter Disc Electronic Digital Scale Model 1036, Tonbridge, UK) and a standard WFR protocol[16]. Recipe data were also collected by weighing the recipe ingredients and final cooked food, and recording the cooking methods (e.g. fried, boiled, stewed). If the child was left in the care of another person, the enumerator remained with the mother, and information about any foods or beverages consumed while the child was away was collected from the secondary caregiver via recall upon their return. The amounts of food/beverage consumed by the mother or child before 06:00 or after 21:00 were recalled and recorded.

On the day after the WFR was collected, two semi-quantitative multiple pass 24-hour dietary recalls were administered to the mother to collect information on all foods and beverages consumed the previous day–one for herself and then one for the child[16]. For each recall, in the first pass, the mother was asked to list everything she (or her child) consumed the previous day; in the second pass, additional details about each food were recorded, including the time of consumption and ingredients in mixed dishes. In the third pass, mothers were asked to confirm the food groups consumed. The quantity of each food consumed was not recorded.

The same day as WFR data collection, a small, lightweight, life-logging wearable camera was attached to a t-shirt worn by the mother at approximately 06:00 and removed at approximately 21:00. Participants were instructed to wear the camera while continuing their usual activities, covering or removing the camera as needed for privacy. The wearable camera automatically recorded a picture every 30-seconds, storing all photos (approximately 1,800) on a memory card.

The following day, an enumerator first reviewed the photos captured by the wearable camera on a tablet and annotated the foods she thought–based on the photos–were consumed by the mother and child i.e. the enumerator image interpretation (EII). The enumerator estimated the dietary diversity score (DDS) for the mother and child based on her interpretation of the photos and demarcated the series of eating episodes for later review with the mother. Upon meeting with the mother, the enumerator first administered the two standard 24HRs (one for the mother and one for her child). Then the enumerator reviewed the photos with the mother on the tablet. During this interview, the enumerator probed the participant based on "the 4 Ws, where appropriate. For example, questions such as: "What… were you doing? Who …were you with? Where …were you? Where…were you going? Where…was the index child? Why …did you go there?" or "Why…were you doing that?". The enumerator revised her original annotations (i.e. the EII) of foods consumed by the mother and child, as needed, based on the mother's feedback. Finally, having reviewed and discussed the previous day's photos with the mother, the enumerator asked the mother to confirm the food groups that she and her child had or had not consumed (i.e. the IAR).

The IAR protocol was adapted from one used in high income country contexts[39]. The protocol followed ethical guidelines for life-logging wearable camera research to ensure that privacy of the participants was maintained[38]. The IAR protocol was pilot tested prior to the start of the study.

Data Processing

The food/beverage recorded over a period of 24 hours were coded into food groups, and a DDS was calculated for each mother and child for each method (WFR, 24HR and IAR). The

DDS for children was based on seven food groups: namely grains, roots and tubers; legumes and nuts; dairy products; flesh foods; eggs; vitamin A-rich fruits and vegetables; and other fruits and vegetables[12]. The DDS for women was based on ten food groups: namely grains, white roots and tubers, and plantains; pulses; nuts and seeds; dairy; meat, poultry and fish; eggs; dark green leafy vegetables; other vitamin-A rich fruits and vegetables; other vegetables; other fruits[13]. The percent of women and children achieving minimum dietary diversity (MDD) was calculated. The threshold MDD used for women was five food groups out of ten; the threshold MDD for children was four food groups out of seven. Breast milk intake was not included in the comparison of the three methods.

Weight-for-age z-score (WAZ), length-for-age z-score (LAZ), and weight-for-length z-score (WLZ) were calculated for each child using the 2006 WHO growth standards[50]; and body mass index (BMI) was calculated for each mother. The proportions of children who were underweight (< -2 SD from median WAZ), stunted (< -2 SD from median LAZ), and wasted (< -2 SD from median WLZ) were calculated. The proportions of mothers who were thin (BMI <18.5), normal (BMI 18.5-24.9), and overweight/obese (BMI ≥25.0) were also calculated. The Ugandan 2012 Poverty Probability Index was calculated, as well as the proportion of the population living below $1.25/day[51].

Data analysis

The primary outcome variables analysed for both mothers and children were DDS and MDD. Data were analysed using Stata/SE version 15.1. P-values less than 0.05 were considered significant for all tests. Cases with incomplete data for any of the three methods (WFR, 24HR or IAR) were eliminated from analysis. A minimum threshold of 13 hours of observation and photos from the wearable camera was deemed adequate; any cases not meeting this threshold were eliminated from analysis to ensure the integrity of the comparison (with the 24HR). The Wilcoxon signed rank sum and McNemar's tests were used to compare the distributions of DDS and MDD, respectively, obtained via the criterion method (WFR) versus the 24HR or IAR. The medians of the DDS differences (24HR minus WFR and IAR minus WFR) were computed, and the distribution of the median DDS differences were also compared. Key socio-demographic characteristics for participating and missing households were compared using the Mann-Whitney two-sample statistic, for continuous data, and the Fisher Exact test for categorical data. DDS and MDD for mothers and children in participating households collected via IAR were also calculated for the non-observation day and compared to those of the corresponding observation day, using the Wilcoxon signed rank sum and McNemar's tests, respectively.

Treating DDS as a continuous measure, the inter-tool agreement between WFR and 24HR or IAR was assessed using the Bland-Altman limits of agreement (LOA) method[46]. Specifically, for each individual, the difference between the methods (DDS estimated using either the 24HR or IAR minus the criterion measure of DDS) versus the mean of the methods were plotted; the relative bias, and the 95% LOA (mean difference ±2 SD of the differences) were estimated. Finally, DDS estimates via the 24HR and IAR methods against the criterion method were also compared using the weighted Cohen's kappa coefficient for inter-rater agreement. It was interpreted as follows: <0.00 Poor agreement; 0.00-0.20 Slight agreement; 0.21-0.40 Fair agreement; 0.41-0.60 Moderate agreement; 0.61-0.80 Substantial agreement; 0.81-1.00 Almost Perfect agreement[52,53].

Results

Characteristics of the sample

Overall, 211 mother-child dyads were recruited into the study. Among those recruited, six participants voluntarily withdrew, and 42 participants were eliminated from analysis due to incomplete data (Figure 2). Characteristics of the study population are presented and compared with participants who were lost to the study in Table 1. These comparisons show there were no differences between participating and missing households, with the exception of child breastfeeding status (61% for participating children vs. 42% for non-participating children). The median household size was six members, and nearly one quarter of participating households lived below $1.25/day. Most participating mothers were married and between the ages of 20 and 29. Nearly two-thirds of participating mothers had not completed primary school, and just under one half were literate. Most mothers were either pregnant, breastfeeding, or both. Most women were of normal BMI.

The median age of participating children was 16.7 months, approximately evenly split between males and females. Nearly all children were initially breastfed, although just 61% were breastfeeding at the time of data collection. Among this population, children were fed by several caregivers in addition to their mother. More than a third of children were fed by at least one caregiver less than 13 years of age. Approximately a quarter of children were stunted but less than 3% of children were wasted.

Diet diversity

The median DDS, for both mothers and children and all methods, was four food groups (Table 2). The estimated percentage of mothers achieving the MDD ranged from 41% for the WFR to 47% for the 24HR; and for children it ranged from 55% for the WFR to 60% for the

24HR (Table 2). The percentage achieving the MDD estimated via the 24HR and IAR, for both mothers and children, was consistently higher than the WFR estimates.

Median DDS and MDD for mothers estimated via IAR on the non-observation day were slightly higher than those collected via IAR on the observation day (5 vs. 4, p=0.2862; and 54% vs. 42%, p=0.1161). (see Supplementary Table 1a and Supplementary Table 1b) For children, median DDS and MDD estimated via IAR on the non-observation day were similar to those collected via IAR on the observation day (4, p=0.5243; and 56% vs. 58%, p=0.3428). (see Supplementary Table 1a and Supplementary Table 1b)

Measure of Agreement

The Bland-Altman plots showed a consistent and uniform pattern across the range of mean DDS for all analyses (Figure 3). The relative bias was low for the 24HR (-0.1801 for mothers; -0.1358 for children;) and the IAR (-0.1227 for mothers; -0.1104 for children;) (Table 3). The percentage of DDS that were identical comparing the IAR or 24HR with the criterion method ranged from 58% (IAR for mothers) to 70% (24HR children). Between 6% and 9% of the estimates erred by 2 or more food groups (see Supplementary Figure 1**)**. Although the relative bias was not clinically important, the LOA were wide (-1.6615 to 1.3012 and -1.6883 to 1.4167 for mothers and children via 24HR, respectively; -2.1322 to 1.8868 and -1.7130 to 1.4921 for mothers and children via IAR, respectively;). For DDS estimated via 24HR and IAR, Cohen's kappa coefficient was 0.68 and 0.59, respectively for the mothers, and 0.60 and 0.59, respectively for the children. (Table 4) For mothers, Cohen's kappa indicated slightly higher inter-method agreement for the 24HR (substantial reliability) than the IAR (moderate reliability), whereas for children the inter-method agreements for the 24HR and IAR are both moderate.

Discussion

This is the first study, to our knowledge, validating the 24HR and IAR method using a life-logging wearable camera for assessing DDS and MDD. Both the 24HR and the IAR provided an accurate estimate of the sample median DDS for women and young children, for the same day of food intake, but tended to overestimate the proportion of women or children that achieved MDD, indicating that the 24HR and the IAR may over-estimate diet quality, at least among women and young children of Eastern Uganda during the dry season.

Although the relative bias seen in this study was low, the high LOA observed for both methods (24HR and IAR) across population groups (mothers and children) was substantial. There are no validation studies of image-assisted recall methods for estimating DDS with which to compare these results. However, similar results have been seen in image-assisted quantitative diet recall validation studies[29,33,35,37,54]. Such error usually serves to attenuate the association between DDS and health or other outcomes and diminish power to detect it[Error! Bookmark not defined.]. This study indicates that attenuation levels might remain high even after accounting for day-to-day variation in DDS, because food groups are misreported by a substantial proportion of individuals.

It is well documented that, in quantitative dietary assessment, measurement error (the difference between reported intake and true intake) commonly occurs[18]. This study suggests that a high degree of measurement error also occurs when diet quality is assessed by the number of selected food groups consumed. In this study, individual estimates of DDS could differ by more than two food groups from observed values. It is remarkable that both 24HR and IAR mis-classified over a third of maternal and child DDS. Although there is no globally accepted threshold LOA for DDS, a difference of one or two food groups (out of 7 for children and 10 for women) is substantial. Errors appear to occur equally at lower- and higher-ends of the DDS spectrum.

The cause of the wide LOAs observed in this study is not immediately evident. Further exploration of the data shows that reported consumption of vitamin A-rich fruits and vegetables was higher (≥5%) in the IAR and 24HR than the WFR for mothers and children, and a higher percentage of mothers reported consuming other fruits (62 vs. 56%) and dairy (22 vs. 17%) in the IAR compared with WFR (see Supplementary Table 2a and Supplementary Table 2b), whereas a lower percentage of mothers reported consumption of other vegetables in the IAR compared with the WFR (90 vs. 95%).

Our finding that the LOA in assessing DDS were high, for both the 24HR and IAR collected for the same day as the criterion method, was somewhat surprising. Compared with the 24HR, we had expected viewing one's own passively collected photographs (IAR) would reduce errors due to memory, social desirability and other biases commonly known to contribute to inaccurate estimates when recalling foods consumed.

There are several plausible contributing factors. For instance, mothers may have become bored and/or fatigued after four similar series of questions about the food groups consumed, thus resulting in more random error in IAR than 24HR, because the mother and child 24HRs were administered before the mother and child IARs. Further review by the author (ALSB) of the data inconsistencies in the procession from WFR to EII to IAR among participants with large discrepancies in DDS suggests that inflation of the DDS from WFR to the IAR may have been due to errors introduced by the mother during the last step of the IAR protocol (i.e. the final confirmation of food groups consumed). In addition to boredom or fatigue, social desirability bias may have contributed to inflation of DDS. No pattern was observed for those that under-estimated the DDS.

Alternatively, and consistent with the long-standing theory that low levels of education and pictorial literacy may affect subjects' capacity to interpret food photos, some women in this study struggled to interpret the first-person photos from the wearable camera[55]. Enumerators were instructed to record foods consumed as reported by the mother, even if it conflicted with their own interpretation of the photos. Thus, error may have occurred, in the IAR, if photos were misinterpreted by the mother.

A thorough analysis of the feasibility and acceptability, using results from the questionnaire administered at the end of the period of data collection and administrative records, will be reported separately (manuscript in preparation). These results may provide further insights into factors contributing to the wide LOAs observed in this study.

Based on the results of this study, the EII (i.e. the enumerator working independently–without the aid of the mother–to interpret from the wearable camera photos foods consumed by the mother or child) did not provide a reliable estimate of DDS for mothers or children. When enumerators annotated foods consumed based on their interpretation of the photos, without the assistance of the mother, they consistently underestimated the variety of foods consumed by both mothers and children compared to the mother-assisted IAR and WFR. (see Supplementary Table 2a and Supplementary Table 2b) For example, based on enumerators' review of photos alone, only 35% of mothers consumed animal source foods compared to 62% when the mother-assisted IAR was used (and 65% in the WFR). For children, only 36% (EII) compared to 60% (IAR) and 59% (WFR) were estimated to have consumed pulses. This suggests that study participants themselves are crucial in interpreting wearable camera images for the purposes of estimating DDS.

DDS and MDD estimated via IAR on observation days were similar to those estimated on non-observation days. The results of this study therefore indicate low reactivity to observation. However, the results achieved in this study may still reflect a higher level of agreement between the 24HR and IAR versus WFR than might otherwise be expected. Owing to higher percentage of breastfeeding children included in the analyses compared to those lost to the study, children in this study may have remained in closer proximity to their mother, thus enabling more consistent monitoring of the child's dietary intake. Also, mothers may have been more vigilant of the child's food intake due to reactivity to the wearable camera.

Overall, dietary quality for mothers and children in this study population was poor. (see Supplementary Table 2a and Supplementary Table 2b) Consumption of most nutrient-dense foods, such as dairy, eggs, and dark leafy greens and other vitamin A-rich fruits and vegetables by mothers and children, respectively, was low. Data collection was conducted during the dry season and, consequently, vitamin A-rich fruit consumption may have been lower than in other seasons. Even though the consumption of animal source foods is relatively high, due to the widespread consumption of small fish, less than half of mothers and less than 60% of children achieved the MDD.

This study was conducted in Bugiri and Kamuli Districts in the Busoga Region of Eastern Uganda where 29% of children under-5 years of age are stunted and 7% of women are underweight[(56)]. In Busoga Region, 66% of women are illiterate, only 12% of women have completed primary school, and more than a third have no regular access to radio, television or the newspaper[(56)]. By comparison, in our study, less than half of the participants were literate and just over one third had only a primary school level of education. Our results show markedly better child diet quality (MDD=55%) compared to that reported by the most recent DHS for the Busoga region (MDD=31%) (DDS was not reported)[(56)]. Relatively higher levels of education and literacy among mothers in this study may be a factor in higher-than-expected child MDD. Participation in the Sasakawa programme, seasonality, method of data collection or secular changes in food consumption patterns may also contribute to differences in prevalence of children having achieved MDD.

Limitations

This study set out to pilot test and evaluate the potential of using an inexpensive life-logging wearable camera to estimate the DDS and MDD of women and young children. Our hypothesis was that prospectively capturing food consumption data would reduce systematic and random errors inherent to dietary recalls and reduce respondent/interviewer burdens inherent to WFRs; and might allow accurate dietary diversity data collection at scale for programmatic purposes in rural LIC contexts. Our results indicate that, although the relative biases of both the 24HR and IAR were low, the high individual-level error observed in both methods may be expected to attenuate associations between DDS and outcomes measured. Therefore, where DDS are estimated via 24HR or IAR data, the true associations between DDS and outcomes may be stronger than they appear as a result of mis-reporting of food group consumption by a large proportion of the population.

In the design of the IAR protocol used in this study, several trade-offs were made. To keep equipment costs low (for LIC contexts), human interaction was required at every step of data processing. For example, although the wearable cameras are fully automated, annotation of the photos (e.g. foods consumed) required for analyses was paper-based and labour-intensive. Humans acting as the bridge between information technology systems provides ample opportunity for error and loss of data.

There were a few common scenarios in which, for the IAR, foods consumed were based solely on recall and were not, in actuality, image-assisted. For example, for reasons of enumerator safety, the wearable cameras often had to be removed before the end of the participants' day, and consequently food preparation and cooking, eating, and feeding activities at the end of the day were often missed. Also, the camera was attached to the mother, so there was no visual record of foods consumed by the child under the care of someone else or when the mother was not facing her child. For this reason, we would have expected a more accurate DDS estimate via IAR for mothers than children. However, this was not observed in any of the key indicators evaluated in this study, indicating other logistical or technical limitations were more important factors contributing to poor agreement. For example, determining ingredients–especially nutrient-dense ingredients commonly consumed in small quantities–posed a serious challenge. In this study, it was rare to see the addition of milk–e.g. into tea or porridge–in photos. Milk was either added by someone other than the mother or otherwise added off camera. Milk and other ingredients may be stored in non-descript, solid-coloured containers, and mothers and children commonly drink from solid-coloured plastic mugs, making it difficult to determine the contents (e.g. to differentiate black tea from milk tea) once served. Differentiating white from yellow tubers, which have different nutrient values, was also a challenge.

In this Ugandan context, food is usually prepared at the family-level and can take hours to cook, with ingredients being added long before consumption, and by various people in the household throughout the day. Also, it was common practice for mothers in this study to prepare food, eat, and feed children while seated on the ground, at an angle such that food consumption was not captured on camera. Food consumption and preparation steps may have also been missed between 30-second photo increments. Fruit, in particular, is often picked and consumed quickly, appearing in just 1 or 2 frames (out of 1,800), or not at all.

Finally, where there is no electricity, pictures taken before dawn, after dark, or inside the kitchen–where a lot of cooking occurs–are too dark to see, and movement of the camera can render photos indecipherable. Addressing these logistical and technical limitations may improve the relative validity of the IAR for estimating DDS and MDD for mothers and children. For children, the IAR may only be an appropriate method for assessing dietary diversity when the caregiver wearing the camera exclusively feeds the child, or for children under 12 months of age who are less mobile and require more assistance during feeding.

Conclusion

The 24HR and IAR performed similarly in estimating maternal and child DDS in this rural LIC context. For both methods and populations, there was low systematic bias. Both 24HR and IAR provided an accurate estimate of median DDS at the population level, although they both tended to overestimate the percentage of mothers and children achieving the MDD. However, importantly, this first-ever study to quantify the extent of measurement error inherent in recall methods for estimating DDS suggests that the degree of attenuation may be greater than previously recognised. Given the high LOA observed here, true associations between DDS–where estimated via 24HR or IAR–and outcomes measured may be stronger than they appear. These results, however, suggest that unless the validity of the IAR can be improved, for reasons of utility, future studies should continue to use data collected using the 24HR to estimate DDS and MDD. The time required for both data collection and processing was substantially lower for the 24HR than the IAR.

Future studies should endeavour to quantify the amount of attenuation due to mis-reporting food group consumption inherent to common methods for assessing DDS, and to investigate factors associated with these errors across different country contexts. As an early prototype tailored to LIC settings, the IAR performed similarly to the 24HR for estimating DDS. Further research and development to address the logistical and technical challenges identified in this study are needed to fully capitalise on the strengths of life-logging wearable cameras for prospectively–and passively–capturing the consumption of food/beverage in a LIC context. Additional studies are needed to determine whether active photography, where participants are instructed to photograph foods when they are consumed and the ingredients added to individual recipes, better addresses the challenges of passive photography identified in this study. Future research should also seek to exploit the unique capability of wearable cameras to simultaneously gather data related to food intake and other factors driving nutrition outcomes (e.g. time allocation, care and feeding practices, availability and accessibility to food, and cleanliness of the environment) to better understand their associations, and inform the design and evaluation of nutrition-sensitive programmes in LICs.

Acknowledgements

This study was financially supported by an Innovative Methods and Metrics for Agriculture and Nutrition Actions (IMMANA) grant, funded by UK AID (J.N., E.L.F., K.W., grant number IMMANA 2.05). IMMANA had no role in the design, analysis or writing of this article. A.L.S.B., L.P., L.F., J.N., K.W., J.P. and E.L.F. contributed to the study design and tool development. A.L.S.B., G.V. and J.N. were involved in data collection and study management. G.V., B.G. and N.S. entered data and contributed to interpretation of the data. A.L.S.B. carried out the validation analyses, with oversight from L.P. and E.L.F. P.K. contributed to the data analysis. A.L.S.B., L.P. and E.L.F. interpreted the results. A.L.S.B. wrote the manuscript. L.P., P.K., L.F., B.G., N.S., J.N., K.W. and E.L.F. edited the manuscript. All authors read and approved the final version of the manuscript. The authors have no conflict of interests to report.

References

1. United Nations. The Millennium Development Goals Report 2010. New York; 2010.
2. UNICEF. Strategy for improved nutrition of infant and women in developing countries. New York; 1990.
3. Black RE, Alderman H, Bhutta ZA, Gillespie S, Haddad L, Horton S, et al. Maternal and child nutrition: building momentum for impact. Lancet. 2013;382(9890):372–5.
4. United Nations Children's Fund (UNICEF), World Health Organization, International Bank for Reconstruction and Development/The World Bank. Levels and trends in child malnutrition: Key Findings of the 2020 Edition of the Joint Child Malnutrition Estimates. Geneva; 2020.
5. Livingstone MBE, Robson PJ, Wallace JMW. Issues in dietary intake assessment of children and adolescents. Br J Nutr. 2004;92(S2):S213–22.
6. Burrows TL, Martin RJ, Collins CE. A Systematic Review of the Validity of Dietary Assessment Methods in Children when Compared with the Method of Doubly Labeled Water. J Am Diet Assoc. 2010;110(10):1501–10.
7. Subar AF, Freedman LS, Tooze JA, Kirkpatrick SI, Boushey C, Neuhouser ML, et al. Addressing Current Criticism Regarding the Value of Self-Report Dietary Data. J Nutr. 2015;145(12):2639–45.
8. Foster E, Bradley J. Methodological considerations and future insights for 24-hour dietary recall assessment in children. Nutr Res. 2018;51:1–11.
9. Working Group on Infant and Young Child Feeding Indicators. Developing and Validating Simple Indicators of Dietary Quality and Energy Intake of Infants and Young Children in Developing Countries: Summary of findings from analysis of 10 data sets. Washington, D.C.: FANTA, FHI 360; 2006.
10. Arimond M, Wiesmann D, Becquey E, Carriquiry A, Daniels MC, Deitchler M, et al. Simple Food Group Diversity Indicators Predict Micronutrient Adequacy of Women's Diets in 5 Diverse, Resource-Poor Settings. J Nutr. 2010 Nov 1;140(11):2059S-2069S.
11. Arimond M, Wiesmann D, Becquey E, Carriquiry A, Daniels M, Deitchler M, et al. Dietary Diversity as a Measure of the Micronutrient Adequacy of Women's Diets in Resource-Poor Areas: Summary of Results from Five Sites. Washington, D.C.: FANTA-2 Bridge, FHI 360; 2011.
12. World Health Organization. Indicators for assessing infant and young child feeding practices: Part 2: Measurement. World Health Organization (WHO); 2010.
13. FAO and FHI 360. Minimum Dietary Diversity for Women: A Guide for

Measurement. Rome: FAO; 2016.

14. Martin-Prével Y, Arimond M, Allemand P, Wiesmann D, Ballard TJ, Deitchler M, et al. Development of a Dichotomous Indicator for Population-Level Assessment of Dietary Diversity in Women of Reproductive Age. Curr Dev Nutr. 2017;1(12):cdn.117.001701.
15. World Health Organization. Global Nutrition Monitoring Framework: operational guidance for tracking progress in meeting targets for 2025. Geneva: World Health Organization; 2017.
16. Gibson RS. Principles of nutritional assessment. 2nd ed. New York: Oxford University Press; 2005. 908 p.
17. Gibson RS, Ferguson EL. An interactive 24-hour recall for assessing the adequacy of iron and zinc intakes in developing countries. Washington, D.C.: International Life Sciences Institute; 2008.
18. Thompson FE, Subar AF. Dietary Assessment Methodology. In: Nutrition in the Prevention and Treatment of Disease. Elsevier; 2017. p. 5–48.
19. Gleason PM, Boushey CJ, Harris JE, Zoellner J. Publishing Nutrition Research: A Review of Multivariate Techniques—Part 3: Data Reduction Methods. J Acad Nutr Diet. 2015 Jul;115(7):1072–82.
20. Fosgate G. Non-differential measurement error does not always bias diagnostic likelihood ratios towards the null. Emerg Themes Epidemiol. 2006 Dec 17;3(1):7.
21. Nelson M, Haraldsdóttir J. Food photographs: practical guidelines I. Design and analysis of studies to validate portion size estimates. Public Health Nutr. 1998 Dec 1;1(4):219–30.
22. Marjan ZM. Evaluation of food photographs for 24-hour recall method. Malays J Nutr. 1995;(1):95–104.
23. Venter CS, MacIntyre UE, Vorster HH. The development and testing of a food portion photograph book for use in an African population. J Hum Nutr Diet. 2000 Jun;13(3):205–18.
24. Turconi G, Guarcello M, Berzolari FG, Carolei A, Bazzano R, Roggi C. An evaluation of a colour food photography atlas as a tool for quantifying food portion size in epidemiological dietary surveys. Eur J Clin Nutr. 2005 Aug 1;59(8):923–31.
25. Robson PJ, Livingstone MBE. An evaluation of food photographs as a tool for quantifying food and nutrient intakes. Public Health Nutr. 2000 Jun 2;3(2):183–92.
26. Faggiano F, Vineis P, Cravanzola D, Pisani P, Xompero G, Riboli E, et al. Validation

of a Method for the Estimation of Food Portion Size. Epidemiology. 1992;3(4):379–82.

27. Small L, Sidora-Arcoleo K, Vaughan L, Creed-Capsel J, Chung K-Y, Stevens C. Validity and Reliability of Photographic Diet Diaries for Assessing Dietary Intake Among Young Children. ICAN Infant, Child, Adolesc Nutr. 2009;1(1):27–36.
28. Flax VL, Thakwalakwa C, Schnefke CH, Stobaugh H, Phuka JC, Coates J, et al. Validation of a digitally displayed photographic food portion-size estimation aid among women in urban and rural Malawi. Public Health Nutr. 2019 Dec 9;22(17):3140–50.
29. Huybregts L, Roberfroid D, Lachat C, Van Camp J, Kolsteren P. Validity of photographs for food portion estimation in a rural West African setting. Public Health Nutr. 2008;11(6):581–7.
30. Harris-Fry H, Paudel P, Karn M, Mishra N, Thakur J, Paudel V, et al. Development and validation of a photographic food atlas for portion size assessment in the southern plains of Nepal. Public Health Nutr. 2016 Oct 21;19(14):2495–507.
31. Amougou N, Cohen E, Mbala ML, Grosdidier B, Bernard JY, Saïd-Mohamed R, et al. Development and validation of two food portion photograph books to assess dietary intake among adults and children in Central Africa. Br J Nutr. 2016 Mar 14;115(5):895–902.
32. De Bruyn J, Msuya J, Ferguson E. Evaluating pictorial charts as a means of collecting participant-recorded data on household dietary diversity in low-literacy communities in Tanzania. Br J Nutr. 2019;122(12):1432–40.
33. Thakwalakwa CM, Kuusipalo HM, Maleta KM, Phuka JC, Ashorn P, Cheung YB. The validity of a structured interactive 24-hour recall in estimating energy and nutrient intakes in 15-month-old rural Malawian children. Matern Child Nutr. 2012 Jul;8(3):380–9.
34. Kaczkowski CH, Jones PJH, Feng J, Bayley HS. Four-Day Multimedia Diet Records Underestimate Energy Needs in Middle-Aged and Elderly Women as Determined by Doubly-Labeled Water. J Nutr. 2000;130(4):802–5.
35. Norman Å, Kjellenberg K, Torres Aréchiga D, Löf M, Patterson E. “Everyone can take photos.” Feasibility and relative validity of phone photography-based assessment of children’s diets – a mixed methods study. Nutr J. 2020 Dec 27;19(1):50.
36. Rollo M, Ash S, Lyons-Wall P, Russell A. Evaluation of a Mobile Phone Image-Based Dietary Assessment Method in Adults with Type 2 Diabetes. Nutrients. 2015 Jun

17;7(6):4897–910.

37. Ashman A, Collins C, Brown L, Rae K, Rollo M. Validation of a Smartphone Image-Based Dietary Assessment Method for Pregnant Women. Nutrients. 2017 Jan 18;9(1):73.

38. Kelly P, Marshall SJ, Badland H, Kerr J, Oliver M, Doherty AR, et al. An Ethical Framework for Automated, Wearable Cameras in Health Behavior Research. Am J Prev Med. 2013;44(3):314–9.

39. Gemming L, Doherty A, Kelly P, Utter J, Ni Mhurchu C. Feasibility of a SenseCam-assisted 24-h recall to reduce under-reporting of energy intake. Eur J Clin Nutr. 2013;67(10):1095–9.

40. Gemming L. Image-assisted dietary assessment: Evaluating the potential of wearable cameras to enhance self -report in the 24-hour dietary recall method. The University of Auckland; 2015.

41. Pettitt C, Liu J, Kwasnicki RM, Yang G-Z, Preston T, Frost G. A pilot study to determine whether using a lightweight, wearable micro-camera improves dietary assessment accuracy and offers information on macronutrients and eating rate. Br J Nutr. 2016;115(01):160–7.

42. Gemming L, Rush E, Maddison R, Doherty A, Gant N, Utter J, et al. Wearable cameras can reduce dietary under-reporting: doubly labelled water validation of a camera-assisted 24 h recall. Br J Nutr. 2015 Jan 28;113(2):284–91.

43. Gemming L, Ni Mhurchu C. Dietary under-reporting: what foods and which meals are typically under-reported? Eur J Clin Nutr. 2016;70(5):640–1.

44. Smith M, Stanley J, Signal L, Barr M, Chambers T, Balina A, et al. Children’s healthy and unhealthy beverage availability, purchase and consumption : A wearable camera study. Appetite. 2019;133(July 2018):240–51.

45. Kamar M, Evans C, Hugh-Jones S. Factors Influencing British Adolescents’ Intake of Whole Grains: A Pilot Feasibility Study Using SenseCam Assisted Interviews. Nutrients. 2019 Nov 1;11(11):2620.

46. Martin Bland J, Altman D. Statistical Methods for Asessing Agreement Between Two Methods of Clinical Measurement. Lancet. 1986 Feb;327(8476):307–10.

47. Giavarina D. Understanding Bland Altman analysis. Biochem Medica. 2015;25(2):141–51.

48. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in medicine: a practical guide. Practical guides to biostatistics and epidemiology. Cambridge ; New

York: Cambridge University Press; 2011. 338 p.

49. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res. 1999;(8):135–60.

50. World Health Organization. The WHO Child Growth Standards. Geneva, Switzerland: World Health Organization (WHO); 2006.

51. Schreiner M. Uganda 2012 Poverty Probability Index (PPI). Saint Louis; 2016.

52. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics. 1977;33(1):159.

53. Farooq M, Doulah A, Parton J, McCrory M, Higgins J, Sazonov E. Validation of Sensor-Based Food Intake Detection by Multicamera Video Observation in an Unconstrained Environment. Nutrients. 2019;11(3):609.

54. Lombard MJ, Steyn NP, Charlton KE, Senekal M. Application and interpretation of multiple statistical tests to evaluate validity of dietary intake assessment methods. Nutr J. 2015 Dec 22;14(1):40.

55. Fuglesang A. Applied Communication in Developing Countries: Ideas and Observations. reprint. Uppsala, Sweden: Dag Hammarskjöld Foundation; 1973. 124 p.

56. Uganda Bureau of Statistics, ICF. Uganda Demographic and Health Survey 2016. Kampala, Uganda and Rockville, Maryland, USA: UBOS and ICF; 2018.

Figures

HHP	DAY 1	DAY 2	DAY 3	DAY 4	DAY 5
1	ICF MQ Anthro	O & I	24HR & IAR	I	IAR TQ
2	ICF MQ Anthro	I	IAR	O & I	24HR & IAR TQ

Fig. 1. Data collection household pattern (HHP). ICF - informed consent form; MQ – mothers' questionnaire; Anthro – anthropometry; O – observation (i.e. weighed food record); I – information and communications technology (e.g. life-logging wearable camera); 24HR – 24-hour recall; IAR – image-assisted recall; TQ – technology questionnaire.

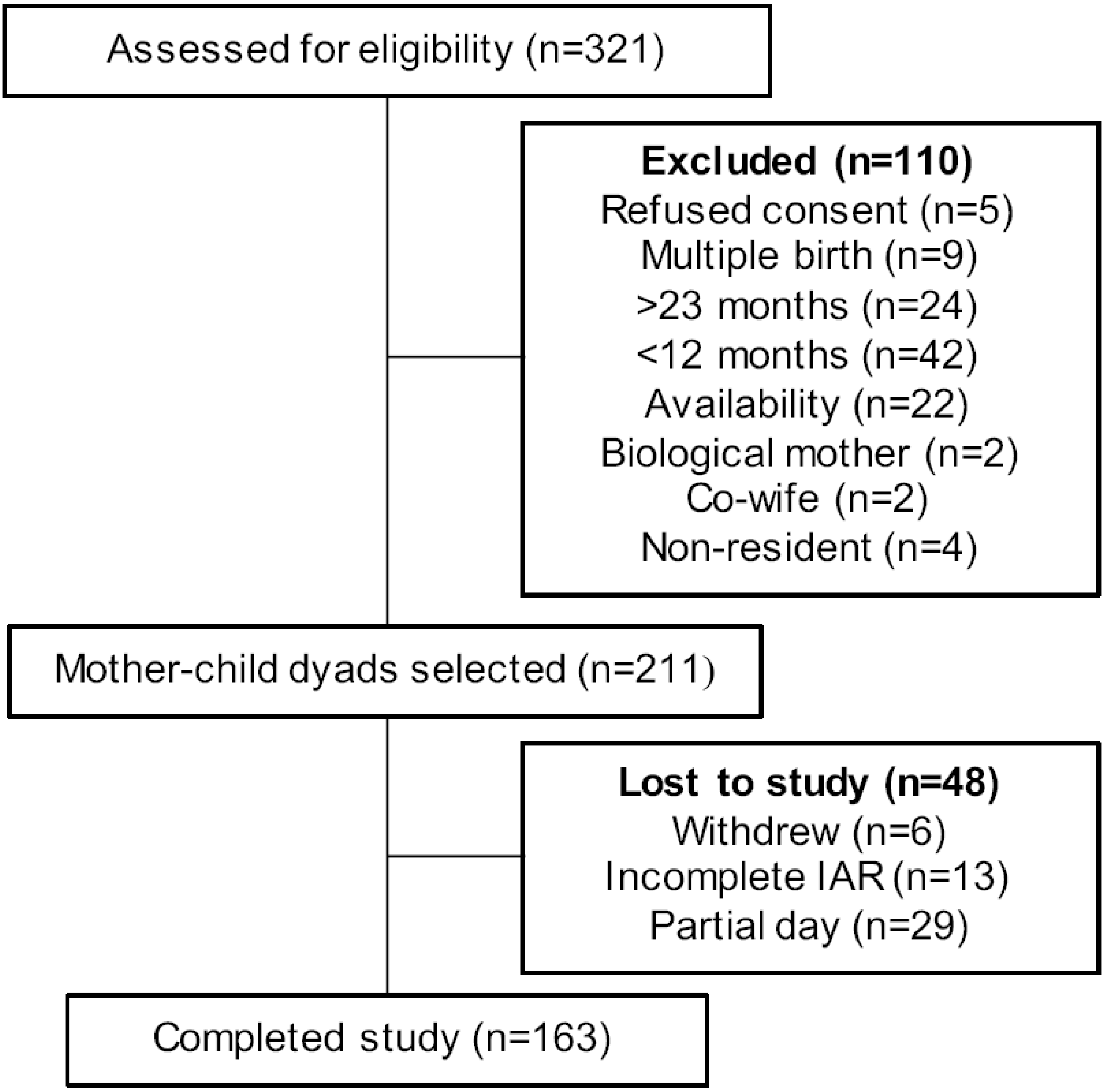


Fig. 2. Study population.

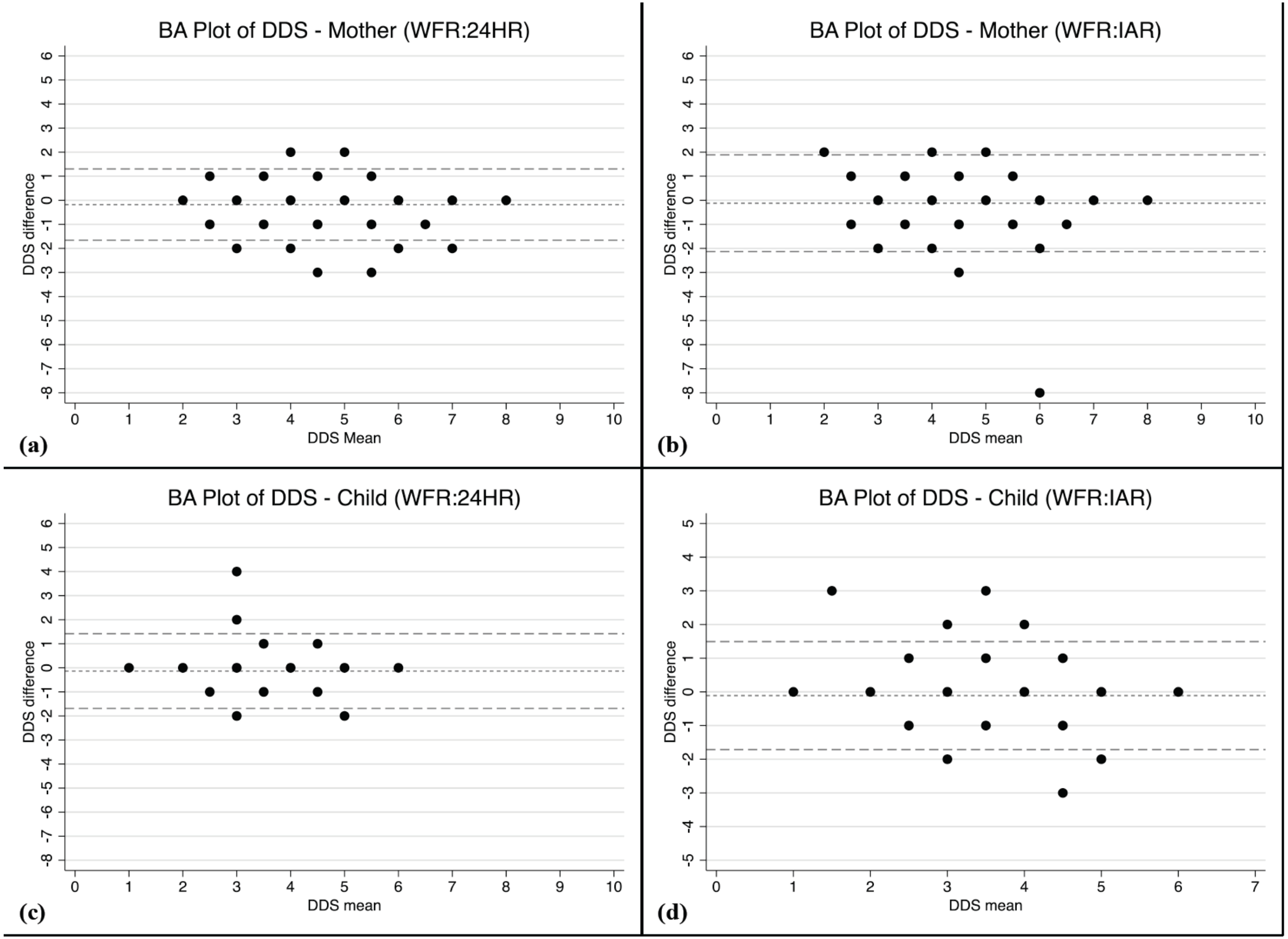


Fig. 3. Bland-Altman (BA) plots of mother and child dietary diversity score (DDS) difference versus the mean of weighed food record (WFR) and 24-hour recall (24HR) or image-assisted recall (IAR). (a) BA plot for mother DDS - WFR and 24HR, (b) BA plot for mother DDS - WFR and IAR, (c) BA plot for child DDS - WFR and 24HR, and (d) BA plot for child DDS - WFR and IAR . The dotted line is relative mean difference (relative bias), the long-dashed lines are +/- 2 SD limits of agreement (LOA).

Tables

Table 1. Characteristics of households, mothers, and children participating in and lost to the study. (Number of participants, proportion of study population, median value, and 25th and 75th percentiles)

	Participating				Missing				
	n	%	Median	25th-75th %	n	%	Median	25th-75th %	*P*
Households									
Number of household members			6	5, 8			5.5	4, 7.5	0.130
Living below $1.25/day (2005 PPP)			20.1	10.9, 27.9			20.1	4, 45.3	0.163
Mothers									
Age (years)			26	22, 30			23	20, 30	0.088
15-19	17	10.4			7	14.9			0.766
20-29	100	61.0			28	59.6			
30-39	39	23.8			11	23.4			
40-49	8	4.9			1	2.1			
Marital status									
Single	20	12.3			1	2.3			0.081
Married or co-habiting	135	82.8			42	95.5			
Level of education									
None or primary incomplete	102	62.2			22	50.0			0.353
Primary complete	55	33.5			19	43.2			
Secondary complete	5	3.1			2	4.6			
Can read and write	78	48.5			24	55.8			0.493
Maternity status									
Pregnant	22	13.9			11	25.0			0.105
Breastfeeding	104	63.4			23	48.9			0.091
Pregnant or breastfeeding	122	74.4			30	63.8			0.196
Body mass index (BMI)			21.7	20.4, 23.9			21.6	19.0, 23.3	0.454
% Thin (BMI <18.5)	7	5.4			4	13.8			0.286
% Overweight/obese (BMI ≥25.0)	17	13.1			3	10.3			
Children									
Age (months)			16.7	14.9, 20.0			17.5	14.7, 19.6	0.682
12-17	99	60.7			24	52.2			0.313
18-23	64	39.3			22	47.8			
Sex									
Female	75	45.7			23	50.0			0.620
Male	89	54.3			23	50.0			
Ever breastfed	161	99.4			42	97.7			0.376
Currently breastfeeding	98	60.5			18	41.9			0.037
Child feeders			3	2, 4			3	2, 4	0.670
No alternative feeders	19	11.6			7	14.9			0.615
All child feeders > 13 years	101	61.6			32	68.1			0.494
Length-for-age z-score (LAZ)			-1.3	-2.1, -0.6			-1.1	-2.2, -0.6	0.939
% below -2 SD	42	25.8			13	29.6			0.508
Weight-for-age z-score (WAZ)			-0.7	-1.5, -0.1			-0.7	-1.5, -0.1	0.848
% below -2 SD	14	9.0			7	18.0			0.067
Weight-for-length z-score (WLZ)			-0.2	-0.8, 0.4			-0.2	-0.6, 0.3	0.776
% below -2 SD	4	2.6			0	0.0			0.678

PPP, purchasing power parity; P, p-value using Mann-Whitney U test to compare the medians and Fisher's Exact test to compare the categorical data.

Table 2. Inter-method comparisons of the median dietary diversity scores (DDS) and percentage achieving minimum dietary diversity (MDD). (median value, and 25th and 75th percentiles)

	Mother						Child					
	WFR (N=163)		24HR (N=161)		IAR (N=163)		WFR (N=163)		24HR (N=162)		IAR (N=163)	
	Median	25th-75th %	Median	25th-75th %	Median	25th-75th %	Median	25th-75th %	Median	25th-75th %	Median	25th-75th %
DDS*	4	3,5	4**	4,5	4	4,5	4	3,4	4**	3,4	4	3,4
	n	%	n	%	n	%	n	%	n	%	n	%
MDD†	67	41.1	76	47.2§	69	42.3	90	55.2	97	59.9	95	58.3

WFR, weighed food record; 24HR, 24-hour recall; IAR, image-assisted recall; DDS, dietary diversity score; MDD, minimum dietary diversity.

* Mother DDS is out of 10 food groups; Child DDS is out of 7 food groups.

† Mother MDD is DDS >= 5; Child MDD is DDS >= 4.

‡ P-value of Wilcoxon signed rank sum test of DDS compared to WFR <0.05.

§ P-value of McNemar's test of the MDD proportion difference compared to WFR <0.05.

Table 3. Inter-method comparisons of the relative bias and limits of agreement (LOA).

	Mother				Child			
	Relative Bias*	LOA†		LOA Difference	Relative Bias*	LOA†		LOA Difference
24HR	-0.1801	-1.6615	1.3012	2.9627	-0.1358	-1.6883	1.4167	3.1050
IAR	-0.1227	-2.1322	1.8868	4.0190	-0.1104	-1.7130	1.4921	3.2051

24HR, 24-hour recall; IAR, image-assisted recall; LOA, limits of agreement.
* Relative mean difference.
† +/- 2 SD from the relative mean difference.

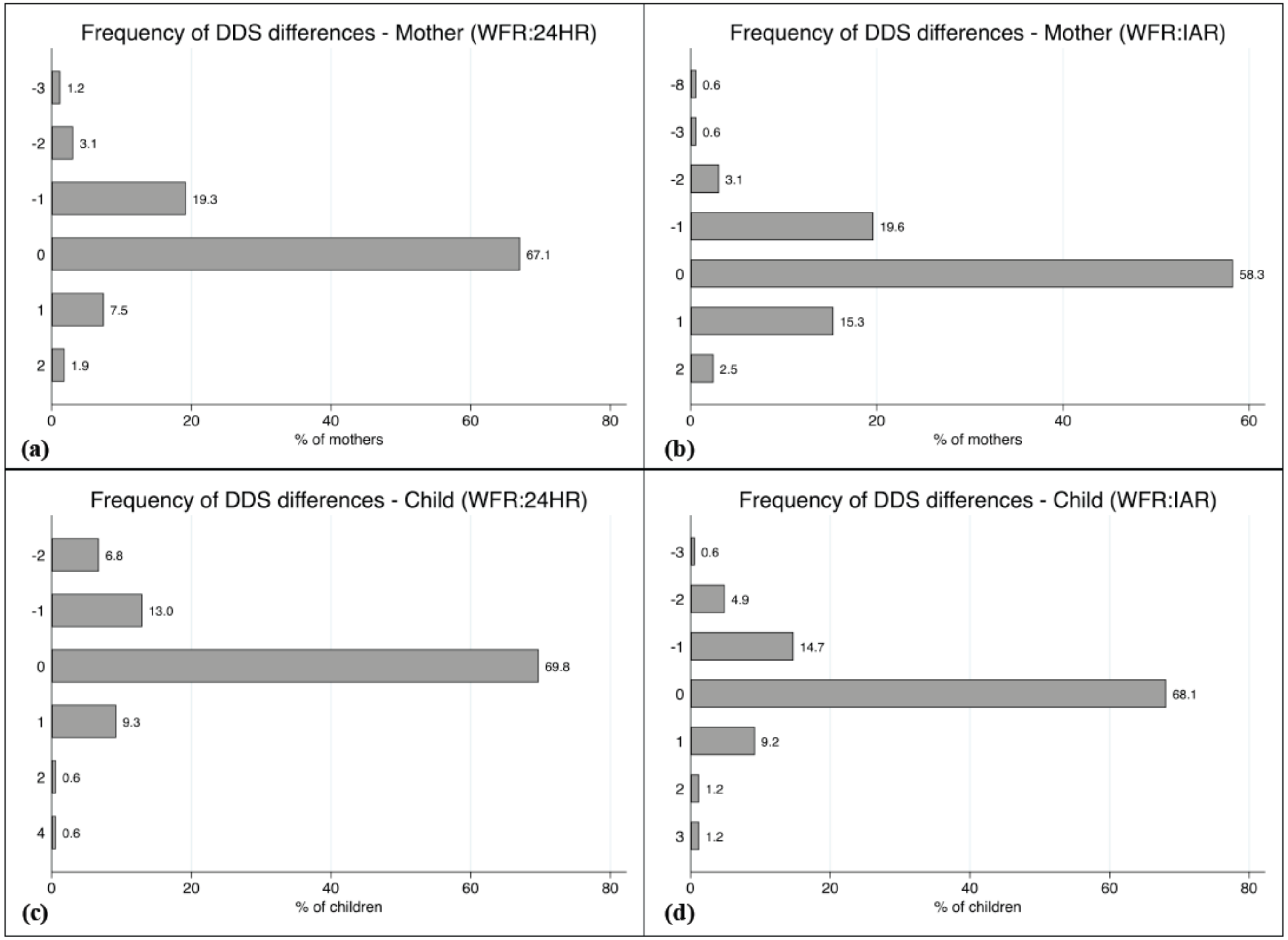
Table 4. Inter-method comparisons of Cohen's kappa coefficient.

	24HR			IAR		
	Cohen's kappa†	95% CI		Cohen's kappa†	95% CI	
Mother*	0.6762	0.5937	0.7587	0.5868	0.4940	0.6796
Child*	0.5989	0.4930	0.7048	0.5945	0.4922	0.6967

24HR, 24-hour recall; IAR, image-assisted recall; CI, confidence interval.
* Mother DDS is out of 10 food groups; Child DDS is out of 7 food groups.
† Using weighted Cohen's kappa for interrater agreement. Landis and Koch (1977) suggest the following benchmark scale for interpreting the kappa statistic: <0.00 Poor; 0.00-0.20 Slight; 0.21-0.40 Fair; 0.41-0.60 Moderate; 0.61-0.80 Substantial; 0.81-1.00 Almost Perfect.

Supplementary Figures

Supplementary Figure 1. Frequency of maternal and child dietary diversity score differences.



Supp. Fig. 1. Frequency of maternal and child diet diversity score (DDS) differences via 24-hour recall (24HR) or image-assisted recall (IAR) versus weighed food record (WFR). (a) Frequency of DDS differences for the mother via 24HR vs. WFR, (b) Frequency of DDS differences for the mother via IAR vs. WFR, (c) Frequency of DDS differences for the child via 24HR vs. WFR, and (d) Frequency of DDS differences for the child via IAR vs. WFR.

Supplementary Tables

Supplementary Table 1a. Inter-method comparisons of the median dietary diversity scores (DDS) for the image-assisted recall (IAR) administered on the same day as the criterion method and on a different day from the criterion method. (median value, and 25th and 75th percentiles)

	IAR		NON OBS IAR		*P*
	Median	25%, 75%	Median	25%, 75%	
Mother*	(N=163)		(N=135)		
	4	4,5	5	4,6	0.2862
Child*	(N=163)		(N=135)		
	4	3,4	4	3,4	0.5243

IAR, image-assisted recall; NON-OBS IAR, non-observation day IAR; P, p-value of Wilcoxon signed rank sum test of NON-OBS IAR versus observation day IAR.
* Mother DDS is out of 10 food groups; Child DDS is out of 7 food groups.

Supplementary Table 1b. Inter-method comparisons of the percentage achieving minimum dietary diversity (MDD) for the image-assisted recall (IAR) administered on the same day as the criterion method and on a different day from the criterion method. (median value, and 25th and 75th percentiles)

	IAR		NON OBS IAR		*P*
	n	%	n	%	
Mother*	(N=163)		(N=135)		
	69	42.3	73	54.1	0.1161
Child*	(N=163)		(N=135)		
	95	58.3	76	56.3	0.3428

IAR, image-assisted recall; NON-OBS IAR, non-observation day IAR; P, p-value of McNemar's test of NON-OBS IAR versus observation day IAR.
* Mother MDD is DDS >= 5; Child MDD is DDS >= 4.

Supplementary Table 2a. Inter-method comparisons of the food groups consumed by the mother. (Number of participants, proportion, mean proportion difference (MD))

	WFR		24HR				EII				IAR			
	n	%	n	%	MD	P	n	%	MD	P	n	%	MD	P
Grains, white roots and tubers, and plantains	163	100.0	162	99.4	0.0061	0.3173	155	95.7	0.0432	0.0082	159	97.6	0.0245	0.0455
Pulses	66	40.5	67	41.1	-0.0061	0.7815	49	30.3	0.1049	0.0031	65	39.9	0.0061	0.7963
Nuts and seeds	53	32.5	55	33.7	-0.0123	0.4142	31	19.1	0.1296	0.0001	56	34.4	-0.0184	0.4054
Dairy	28	17.2	34	20.9	-0.0368	0.0578	11	6.8	0.1049	0.0002	36	22.1	-0.0491	0.0209
Meat, poultry and fish	106	65.0	105	64.4	0.0061	0.7055	56	34.6	0.3025	<0.001	100	61.4	0.0368	0.1088
Eggs	5	3.1	5	3.1	0.0000	1.0000	0	0.0	0.0252	0.0455	8	4.9	-0.0184	0.0833
Dark green leafy vegetables	13	8.0	14	8.6	-0.0061	0.3173	10	6.2	0.0185	0.3173	15	9.2	-0.0123	0.1573
Other vitamin A-rich fruits and vegetables	21	12.9	31	19.0	-0.0613	0.0124	17	10.5	0.0247	0.4142	34	20.9	-0.0798	0.0046
Other vegetables	155	95.1	157	96.3	-0.0123	0.3173	74	45.7	0.4938	<0.001	147	90.2	0.0491	0.0325
Other fruits	91	55.8	100	61.7	-0.0556	0.0290	85	52.8	0.0248	0.5050	101	62.0	-0.0613	0.0330

WFR, weighed food record; 24HR, 24-hour recall; EII, enumerator image interpretation; IAR, image-assisted recall; MD, mean proportion difference (versus WFR); P, p-value of McNemar's test of the proportion difference compared to WFR.

Supplementary Table 2b. Inter-method comparisons of the food groups consumed by the child. (Number of participants, proportion, mean proportion difference (MD))

	WFR		24HR				EII				IAR			
	N	%	N	%	MD	P	N	%	MD	P	N	%	MD	P
Grains, white roots and tubers, and plantains	163	100.0	160	98.2	0.0184	0.0833	149	92.0	0.0802	0.0003	161	98.8	0.0123	0.1573
Pulses	96	58.9	99	60.7	-0.0184	0.4913	59	36.4	0.2284	<0.001	98	60.1	-0.0123	0.6374
Dairy	38	23.3	42	25.8	-0.0245	0.2059	14	8.6	0.1481	<0.001	43	26.4	-0.0307	0.0956
Meat, poultry and fish	97	59.5	100	61.4	-0.0184	0.3657	52	32.1	0.2778	<0.001	98	60.1	-0.0061	0.8185
Eggs	6	3.7	6	3.7	0.0000	1.0000	1	0.6	0.0248	0.0455	9	5.5	-0.0184	0.0833
Other vitamin A-rich fruits and vegetables	32	19.6	46	28.2	-0.0859	0.0010	18	11.1	0.0864	0.0060	43	26.4	-0.0675	0.0218
Other fruits or vegetables	157	96.3	158	96.9	-0.0061	0.6547	112	69.1	0.2716	<0.001	155	95.1	0.0123	0.4795

WFR, weighed food record; 24HR, 24-hour recall; EII, enumerator image interpretation; IAR, image-assisted recall; MD, mean proportion difference (versus WFR); P, p-value of McNemar's test of the proportion difference compared to WFR.